Multi-Topographic Neural Network Communication and Generalization for Multi-Viewpoint Analysis

Shadi Al Shehabi, Jean-Charles Lamirel

▶ To cite this version:

Shadi Al Shehabi, Jean-Charles Lamirel. Multi-Topographic Neural Network Communication and Generalization for Multi-Viewpoint Analysis. International Joint Conference on Neural Networks - IJCNN'05, Jul 2005, Montréal/Canada, pp.1564–1569. inria-00000842

HAL Id: inria-00000842

https://inria.hal.science/inria-00000842

Submitted on 29 Nov 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multi-Topographic Neural Network Communication and Generalization for Multi-Viewpoint Analysis

Shadi Al Shehabi
LORIA
Campus Scientifique, BP 239
54500 Vandoeuvre-lès-Nancy Cedex
France
E-mail: Shadi.Al-Shehabi@loria.fr

Jean-Charles Lamirel
LORIA
Campus Scientifique, BP 239
54506 Vandoeuvre-lès-Nancy Cedex
France
E-mail: Jean-Charles.Lamirel@loria.fr

Abstract— **This paper presents a new generic multi-topographic neural network model whose main area of application is clustering and knowledge extraction tasks on documentary data. The most powefull features of this model are its generalization mechanism and its mechanism of communication between topographies. This paper shows how these mechanisms can be exploited within the framework of the SOM and NG models. An evaluation of the generalization mechanism based on original quality and propagation coherency measures is also proposed. A secondary result of this evaluation is to proof that the generalization mechanism could significantly reduce the well-known border effect of the SOM map.**

I. Introduction

The Kohonen self-organizing map (SOM) model is a specific kind of ANN which implements in only one step the tasks of clustering and mapping a data set. In the SOM case, learning is competitive and unsupervised and the approach gives central attention to spatial order in the clustering of data. The purpose is to compress information by forming reduced representations of the most relevant features, without loss of information about their interrelationships. In the quantitative studies of science, the SOM model has been successfully used for mapping scientific journal networks [1], or author co-citation data [2]. The accuracy of the SOM model has also been demonstrated in the general field of data analysis as well as for documentary database contents mapping and browsing [3]. The MultiSOM model has been proposed in [4]. It represents a significant extension of the SOM model. The principle of this model is to use multiple viewpoints, each one being represented by a single SOM map, in order to enhance both the quality and the granularity of the data analysis and to reduce the noise which is inevitably generated in an overall classification approach. The conservation of an overall view of the analysis is achieved through the use of a communication mechanism between the maps. The advantage of the multi-viewpoint analysis provided by MultiSOM as compared to the global analysis provided by SOM has been clearly demonstrated for precise mining tasks such as patent analysis [5] or webometrics [6]. Another important mechanism provided by the MultiSOM model is the generalization mechanism. This mechanism consists in starting from the original map and introducing new classification levels of synthesis (i.e. maps) by progressively reducing the number of neurons. One of its main advantages is that it does not necessitate any new learning phase. The original role of the generalization mechanism was to highlight to the analyst both the most stable and the most generic results of the analysis [7]. New experiments have shown that this mechanism can also perform error correction in the context of the SOM model. In order to enhance the quality of multi-viewpoint analysis, we propose here to use a neural gas (NG) model as a basis for extending the MultiSOM model to a MultiGAS model. Hence, the NG model is known as more efficient than the SOM model for classification tasks where explicit visualization of the data analysis results is not required. This paper illustrates the respective behavior of the inter-communication and generalization mechanisms within the framework of the SOM and NG models. The first section of the paper presents the two unsupervised neural methods, SOM and NG. The second section presents the MutiSOM model peculiarities (inter-map communication and generalization mechanisms) along with their extension to the MultiGAS model. The last section provides an evaluation of the generalization mechanism using original quality and propagation coherency measures.

II. Self-Organizing Map and Neural Gas

The architecture form of the SOM network is based on the understanding that the representation of data features might assume the form of a self-organizing feature map which is geometrically organized as a grid. A mapping from a high-dimensional data space R^n onto a two dimensional grid of neurons is thus defined. The SOM algorithm is presented in details in ([8]). It consists of two basic procedures: (1) selecting a winning neuron on the grid and (2) updating weights of the winning neuron and of its neighboring neurons. Once the SOM algorithm is achieved, the training data can be affected to the neurons of the map. A SOM map construction is not a straightforward process. It necessitates several different learning steps, single map evaluations, and comparisons between a lot of generated maps in order to find at least a reliable map, at the best an optimal one [4],

[6]. Moreover, special care must be taken of a well-known problem related to the SOM trained structure, namely the border effect. It means that units on edges of the network do not stretch out as much as they should towards the outliers data [8], Last but not least, the neurons of SOM do not necessarily get close to the structure of the data because of the fixed topological structure of the grid.

In the NG algorithm (see [9]), the weights of the neurons are adapted without any fixed topological arrangement within the neural network. Instead, this algorithm utilizes a neighborhood ranking process of the neuron weights for a given input data. The weight changes are not determined by the relative distances between the neuron within a topologically pre-structured lattice, but by the relative distance between the neurons within the input space, hence the name ”neural gas” network. Indeed, thanks to the loss of topographic constraints as compared to SOM, NG tends to better represent the structure of the data, yielding better classification results [10].

III. Multi-Topographic model

The communication between self-organizing maps has been firstly introduced in the context information retrieval for analyzing the relevance user's queries regarding the documentary database contents [4]. It represents a major amelioration of the basic Kohonen SOM model. From a practical point of view, the MultiSOM model introduces the use of viewpoints in the information analysis. In documentary data analysis, the viewpoint building principle consists in separating the description of the documents into several subdescribtions corresponding different keyword subsets. These subsets may fit into the structure of the document when they correspond to different index vocabulary subsets associated to the different document sub-fields. Specific viewpoints may be associated to specific reference fields like ”indexer keywords”, ”title keywords”, or ”author” field. Complementary viewpoints may be also extracted from the overall document description space. In the MultiSOM model, each analysis concerning a given viewpoint is carried out in the form of a map. Each single map is itself a spatial order in which the information is represented both into neurons (classes or topics) and spatial areas (group of classes or macro-topics) [4], [3].

A. Inter-map communication mechanism

The inter-map communication mechanism makes it possible to highlight (both in an automatic or in an interactive way) semantic relationships between different topics belonging to different viewpoints. In MultiSOM, this communication is based on the use of the data that have been projected onto the maps as intermediary neurons or activity transmitters between maps. The intercommunication process between maps operates in three successive steps. Fig. 1 shows graphically the three steps of this intercommunication mechanism. At step 1, the original activity is directly set up on a neuron or on a logical area of a source map through different scalable modalities (full acceptance, moderated acceptance, moderated rejection, full rejection) directly associated to neurons activity levels. The role of this procedure is to highlight (positively or negatively) different topics representing potential centers of interest relatively to the viewpoint associated to the source map.

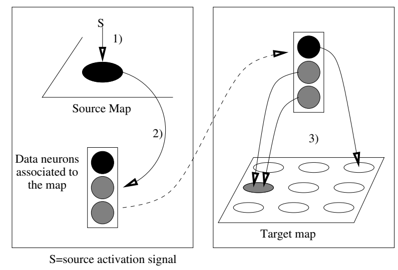

Fig. 1. Inter-map communication mechanism

The activity transmission to target maps is based itself on two elementary steps: a first transmission step from the activated source map to its associated data neurons (down activation), and a second transmission step from the activated data neurons to the target map (up reactivation). The inter-map communication is established by a standard Bayesian inference network propagation algorithm, used for computing the posterior probabilities of target map's neurons T_k which inherited of the activity (evidence Q) transmitted by its associated data neurons D. These computations can be carried out efficiently thanks to the specific Bayesian inference network topology that can be associated to the MultiSOM model. Hence, it is possible to compute the probability $P(act|T_k, Q)$ for an activity of modality act_m on the map neuron T_k which is inherited from activities generated on the source map. This computation is achieved as follows [11]:

$$P(act_m|T_k, Q) = \frac{\sum_{D_j \in act_m, T_k} Sim(D_j, S_i)}{\sum_{D_j \in T_k} Sim(D_j, S_i)} \quad (1)$$

$$P(Act|T_k, Q) = Argmax_i\{P(act_i|T_k, Q)\} \quad (2)$$

The belonging degree of a data Di to a map neuron S_i is computed with the cosine correlation measure:

$$Sim(D_i, S_i) = \frac{D_i \bullet S_i}{\|D_i\|.\|S_i\|} \quad (3)$$

where $\|D_i\|$ represents the norm of the index vector associated to the data D_i, $\|S_i\|$ the norm of the codebook vector associated to the neuron S_i, and $\bullet$ represents the scalar product.

The neurons of the target maps getting the highest probabilities as regard to the equation (3) can be considered as the ones who include the topics sharing the strongest relationship with the topics belonging to the activated neurons of the source map. As soon as the activity transmission has been performed through the projected data, the MultiSOM inter-communication process is generic enough to be directly applied to any other model managing viewpoints. In particular, it can be applied without change to a MultiGAS model. In such a model, each single viewpoint will then be represented by a specific gas.

B. Generalization mechanism

The main objectives of the generalization method are to evaluate the coherency of the topics that have been computed on an original map and to summarize the contents of this later into more generic topics.

1) Generalization of SOM: Let $n \times m$, $(n, m \geq 2)$ be the dimension of the map associated to a given level, the hereabove described generalization process will then produce a next more general level in the form of a $(n-1) \times (m-1)$ map (Fig. 2). For each new level neuron n the following codebook vector computation applies:

$$W_n^{M+1} = \frac{1}{4} \sum_{n_k \in V_n^M} W_{n_k} \tag{4}$$

where V_n^M represents the square neighborhood set on the map M associated to the neuron n of the new synthetic map $M+1$.

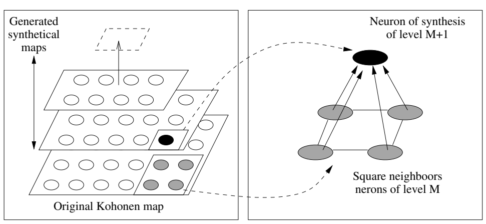

Fig. 2. Map generalization Mechanism

2) Generalization of NG: In the case of the NG model, the original connections between the neurons, which are created by the Neural Gas plus Competitive Hebbian Learning method [12] do not determine neighborhood relations of the same kind as the SOM model. Thus, these connections can not be used in the generalization mechanism because they do not reflect any fixed topology. Moreover, they are specifically instable because they depend on an empirical connection age parameter. Our NG generalization mechanism solves this problem by creating its specific link structure in which each neuron of a given level is linked to its 2-nearest neighbors (Fig. 3). For each new level neuron n the following codebook vector computation applies:

$$W_n^{M+1} = \frac{1}{3}(W_n^M + \sum_{n_k \in V_n^M} W_{n_k}) \tag{5}$$

where V_n^M represents the 2-nearest neighbor neurons of the neuron n on the level M associated to the neuron n of the new generated level $M+1$. After the codebook vector computation the repeated neurons of the new level (i.e. the neurons of the new level that share the same codebook vector) are summarized into a single neuron.

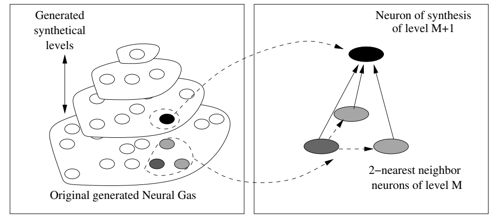

Fig. 3. Gas generalization Mechanism (2D representation of gas is used for the sake of clarity of the figure)

Existing clustering algorithm, such as growing hierarchical self-organizing map (GHSOM) [13], represents a dynamically growing architecture which evolves into a descending hierarchical structure of SOM [14]. Nevertheless, the weak point of this method is to isolate lower level maps without regards to their potential links. As opposed, our generalization method has the advantage of preserving the original neighborhood structure on the new generated levels. Moreover, it ensures the conservation of topographic properties of the map neuron codebook vectors [4]. It could be also considered as an implicit and distributed form of a hierarchical classification method based on neighborhood reciprocity [7]. Finally, there is a straightforward relationship between the generalization and the inter-map communication mechanism. Therefore, the inter-map communication could be used between map/gas and its generalizations in so far as they share the same projected data.

IV. Evaluation of the generalization mechanism

A. Evaluation criteria

In order to evaluate the generalization mechanism that we have defined in the context of the MultiSOM and the MultiGAS models, two kinds of evaluation measures will be used. The first kind will be our own quality criteria, that is to say the Precision and Recall measures based on the properties of class members defined in [5]. As opposed to classical intra-class and inter-class inertia measures, our measures have the main advantage to be independent of the classification methods [6]. The Precision criterium measures in which proportion the contents of the classes generated by a classification method is homogeneous. The greater the Precision, the nearer the intentions of the data belonging to the same classes will be one with respect to the other. As a final result, the classes will be more homogeneous. In a complementary way, the Recall criterium measures the exhaustiveness of the contents of said classes, evaluating to what extent single properties are associated with single classes. These measures can be firstly applied to estimate the overall quality of a classification. In a complementary way, the break-even points between Recall and Precision can be used to determine an optimal number of classes for a classification related to a given dataset. In the context of the comparision of a basic classification with its generalizations, the expected results will be the reduction of the Precision values and the increase of the Recall values when going up through the generalization levels.

The second kind of evaluation measure is the propagation coherency measure which is carried out with our Bayesian propagation model (see previous section). The coherency measure will evaluate the activity focalization generated by a source map on a target map (Fig. 4). The activity propagation from a source map to a target map is defined by the function:

$$PRG : S_k \rightarrow T_{i_k}$$

The propagation coherency (PC) is then given by:

$$PC = \frac{1}{\overline{S}} \sum_{k; S_k \neq \emptyset} \frac{\sum_i P(act|T_{i_k})}{D_k + 1} \tag{6}$$

$$D_k = \frac{2 \sum_i \sum_{j=i+1} \|T_{i_k} - T_{j_k}\|}{|T^k|.(|T^k| - 1)}$$

$$T_i = \begin{cases} (w_1^i, ..., w_n^i) & \text{NG: original vector space position} \\ (a^i, b^i) & \text{SOM: coordinate position on the grid} \end{cases}$$

$\overline{S}$ represents the peculiar set of neurons extracted from the source map of S, which verifies:

$$\overline{S} = \{S_k \in S | S_k \neq \emptyset, S_k \in act\}$$

$$T_{i_k} \in PRG(S_k)$$

$$T^k = \{T_{i_k} \in T | T_{i_k} \in PRG(S_k)\}$$

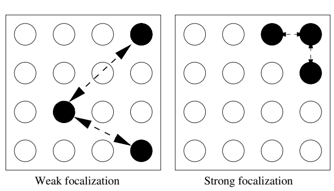

Fig. 4. Activity profiles on target maps

This measure can be used to evaluate the topologic coherence between two neural classifications. Two measures will be necessary to compare an original classification with one of its generalization levels: an up-coherency measure and a down-coherency measure. In the up-coherency measure, the original classification is considered as the source map/gas and its generalization level as the target map/gas. This measure evaluates the quality of the generalization mechanism. The higher is the up-coherency measure, the better will be the generalization mechanism. Indeed, even the number of neurons of the generalization level is smaller than in the original level, a good generalization mechanism will not spread the original data from the original map/gas into the generalized level. In the down-coherency measure, the original classification is considered as the target map/gas and its generalization level as the source map/gas. This measure evaluates the discrimination power of the original classification as compared to its generalization. The higher the down-coherency measure, the less discriminant will be the original classification as compared to its generalization level. Indeed, an original classification will not be discriminant as compared to its generalization, if it only uses the same number of neurons than the latter for projecting the original data.

B. Experimental data

Our experiments consist in evaluating the performance of both generalization mechanisms as described in the previous sections. Our test database contains 1000 patents that has been used in some of our preceding experiments [5]. For the viewpoint-oriented approach the structure of the patents has been parsed in order to extract four different subfields corresponding to four different viewpoints: Use, Advantages, Titles and Patentees. As it is full text, the content of the textual fields of the patents associated with the different viewpoints is parsed by a lexicographic analyzer in order to extract viewpoint specific indexes. For each specific viewpoint the resulting index set is weighted by means of an IDF weighting scheme [15]. Only a single USE viewpoint will be considered in our experiment. This viewpoint generates itself a description space of size 234. Each of our experiments is initiated with an optimal map/gas generated thanks to an optimization algorithm based on the quality criteria [6].

Our first experiment consists in generating a set of maps for the MultiSOM model:

- First, original maps of 16×16 (optimal, see section IV.A), 15×15, 14×14 and 13×13 neurons are generated.
- Generalized maps of 15×15, 14×14 and 13×13 neurons are generated by respectively applying the generalization mechanism to the 16×16 (optimal), 15×15, 14×14 original maps.
- Furthermore, six levels of embeded generalization (from 15×15 neurons to 10×10 neurons) were applied to the original optimal map of 16×16 neurons.

Our second experiment consists in generating a set of gases for the MultiGAS model:

- First, original gases of 121 (optimal, see section IV.A) and 81 are generated.
- Generalized gases of 98 and 64 neurons are generated by respectively applying the generalization mechanism to the 121 and 81 original gases.
- A new original gas of 64 neurons is generated in a third step (same number of neurons as the generalized one).
- In a last step, a generalized gas of 53 neurons is generated by applying the generalization mechanism to the 64 original gas.

For each model, the propagation coherency between each original map/gas and its direct original neighbor of lower class count is computed, as well as the propagation coherency between each original map/gas and its direct generalized map/gas (Fig. 6, Fig. 9). Moreover, the quality measures are computed for each map/gas whether they be original or generalized (Fig. 5, Fig. 8). For the MultiSOM model, the quality of different levels of generalized map is computed as well (Fig. 7).

C. Results

As a first result, the comparison between the quality measures of the optimal NG gas of 121 neurons and the optimal SOM map of 256 neurons shows clearly the advantage of the NG model as compared to the SOM model. Hence, for a number of neurons which is almost twice lower, the NG optimal gas Recall and Precision values are higher than those of the SOM optimal map (Fig. 5, Fig. 8). The higher homogeneity of the results of the NG model as compared with those of the SOM model is also highlighted by the up-coherency values between original maps/gases. Indeed, the average up-coherency between original gases of different neuron number is higher than the average up-coherency between original maps of different neuron number, even when these maps/gases have been generated using the same data (Fig. 6, Fig. 9). In the case of both models, Fig. 6 and Fig. 9 show that the up-coherency values between original maps/gases and their generalized maps/gases are high. Moreover, these values are significantly higher than the up-coherencies values between original maps/gases described in the latter section. This results prove the accuracy of the generalization process for both models. In the case of the MultiSOM model, the systematically lower precision values (for similar value of recall) of original maps as compared to their generalized maps could be considered as a paradoxical result (Fig. 5) (see section IV.A for expected results). Indeed, this indicates that the quality of the generalized maps is higher than that of the original related maps. It also means that the mechanism of generalization makes it possible to significantly reduce the well-known border effect of the original SOM map through the first level of its generalization. When the number of outliers is sufficiently high, like in documentary data, our hypothesis is that all the codebook vectors of the map become sensitive to the border effect. In a first step our generalization mechanism will then provide an adjustement of the codebook vectors instead of performing an efficient generalization. The hypothesis of codebook vector adjustement in the generalized maps can be confirmed by the fact that the down-coherency values between the original maps and their generalized maps are significantly higher than those between original maps (Fig. 6). Indeed, this last result indicates that the map and its generalization tend to become similar. Even if the neuron number of an original SOM map is higher than the one of its associated generalized map, the later number will be used in both maps for effective data representation. This phenonemon is not verified with the MultiGAS model (Fig. 9). On its own side, the generalization mechanism provided by the MultiGAS model directly behaves in a suitable way by reducing the Precision value and increasing the Recall value from the first level of generalization (Fig. 8). The examination of Fig. 7 shows that the generalization mechanism provided by the MultiSOM model will also behave as expected as soon as it goes up to the second level of generalization. Lastly, the results of the generalization mechanism are slightly better in the MultiSOM model than in the MultiGAS model (better up-coherency values on average for the generalization with MultiSOM model) (Fig. 6, Fig. 9). In the N-dimensional space of NG, neighborhood links generated between neurons of a gas could be shared by many generated triangles, leading to potential inaccuracy in data projection on the gas generalization. The mechanism of the MultiSOM model which generalize the original maps from a single square neighborhood set of neurons should be logically less sensitive to this phenomenon.

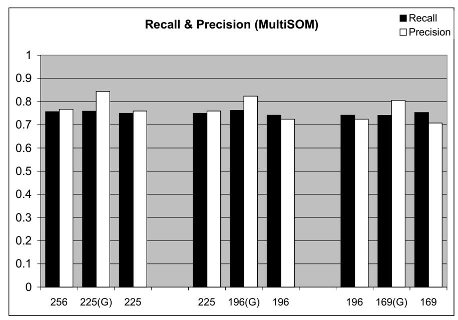

Fig. 5. MultiSOM model: Evaluation of the Quality of original and generalized maps for Use viewpoint (x(G): represents the first level of a generalized map of x neurons)

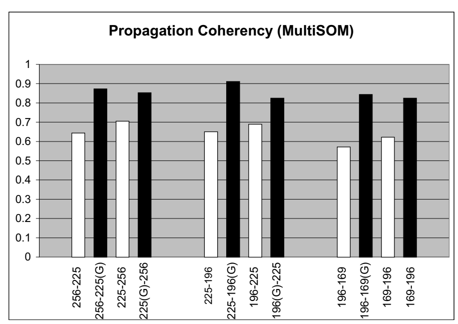

Fig. 6. MultiSOM model: Evaluation of the Propagation Coherency for the Use viewpoint (x-y: represents the propagation from the source map x to a destination map y)

V. Conclusion

In this paper we have proposed both an extension of the inter-communication and the generalization mechanisms of the MultiSOM model to a MultiGAS model as well as an overall evaluation of the generalization mechanism for the two models. The evaluation of the generalization mechanism has been performed using original quality and propagation

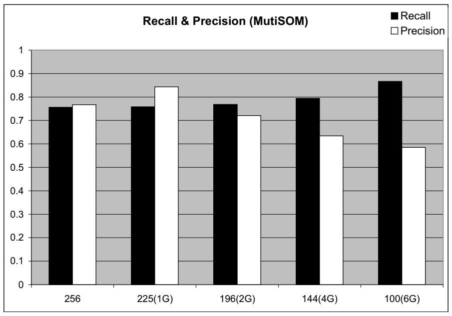

Fig. 7. MultiSOM model: Evaluation of the Quality of the levels of generalization for Use viewpoint (xG: represents the x^{th} level of map generalization)

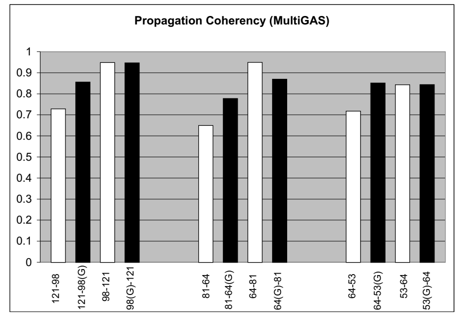

Fig. 9. MultiGAS model: Evaluation of the Propagation Coherency for the Use viewpoint

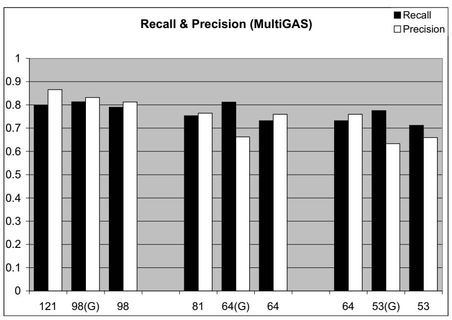

Fig. 8. MultiGAS model: Evaluation of the Quality of original and generalized gases for Use viewpoint

coherency measures. Our experimental results have proven that our generalization mechanism is an efficient process, both for the MultiSOM and the MultiGAS models. An unexpected result of the evaluation of the generalization mechanism provided by the MultiSOM model is that this mechanism permits to significantly reduce the well-known border effect of the original SOM map through the first level of its generalization. New experiments have shown that the generalization mechanism can perform unsupervised knowledge extraction tasks, like rule extraction, when used in combination with the inter-map communication mechanism. As compared to classical symbolic rule extraction methods, like Galois lattice, it has the advantage of performing more accurate rule selection. As soon as for knowledge extraction tasks explicit visualization of the data analysis results is not required, a suitable strategy should be to use our new MultiGAS model instead of our former MultiSOM model for these tasks. Indeed, the experiments described in this paper have also highlited the higher global accuracy and homogeneity of the MultiGAS model as compared to the MultiSOM model.

REFERENCES

[1] J. M. campanario, "Using neural networks to study networks of scientific journals," *Scientometrics*, vol. 33, no. 1, pp. 23–40, 1995.

[2] H. D. White, "Two modes of automated domain analysis: Multidimensional scaling vs kohonen feature mapping of information science author," in *Proceedings of the Fifth International ISKO Conference, Lille, France.*, August 1998, pp. 57–61.

[3] X. Lin, D.Soergel, and G. Marchionini, "A self organizing semantic map for information retrieval," in *Proceedings of 4th International SIGFIR Conference on R&D in Information retrieval*, 1991, pp. 262–269.

[4] J.-C. Lamirel, "Application d'une approche symbolico-connexioniste pour la conception d'un systm̀e documentaire hautement interactif," Ph.D. dissertation, Université de Nancy 1 Henri Poincaré, Nancy, France, 1995.

[5] J.-C. Lamirel, S. Al Shehabi, M. Hoffmann, and C. Francois, "Intelligent patent analysis through the use of a neural network : experiment of multi-viewpoint analysis with the multisom model," in *The ACL 2003 Workshop on Patent Corpus Processing, Sapporo, Japan*, Jul 2003, pp. 7–23.

[6] J.-C. Lamirel, S. Al Shehabi, C. Francois, and M. Hoffmann, "New classification quality estimators for analysis of documentary information : application to patent analysis and web mapping," *Scientometrics*, vol. 60, no. 3, pp. 445–462, Feb 2004.

[7] J.-C. Lamirel, J. Ducloy, and G. Oster, "Adaptive browsing for information discovery in an iconographic context," in *RIAO'2000, Collège de France, Paris, France*, CID, Ed. CID, Apr 2000.

[8] T. Kohonen, *Self-Organising Maps*, 3rd ed. Berlin: Springer-Verlag, 2001.

[9] Martinetz, Thomas and Schulten, Klaus, "A "Neural-Gas" Network Learns Topologies," in *Artificial Neural Networks – Proceedings of the 1991 International Conference on Artificial Neural Networks (ICANN'91)*, Kohonen, T., Mäkisara, K., Simula, O., and Kangas, J., Eds., vol. 1. Elsevier Science Publishing Company B. V., 1991, pp. 397–402.

[10] S. Al Shehabi and J.-C. Lamirel, "Unsupervised neural networks of topographic and gas families for documentary data classification," in *The 8th World Multi-Conference on Systemics, Cybernetics and Informatics - SCI'04, Orlando, Florida, USA*, Jul 2004.

[11] J.-C. Lamirel, S. Al Shehabi, C. Francois, and X. Polanco, "Using a compound approach based on elaborated neural network for webometrics : An example issued from the eicstes project," *Scientometrics*, vol. 61, no. 3, pp. 427–441, Jan 2004.

[12] B. Fritzke, "Some competitive learning methods," Ruhr-Universitat Bochum, Tech. Rep., 1997.

[13] M. Dittenbach, D. Merkl, and A. Rauber, "The growing hierarchical self-organising map," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN2000)*, vol. VI, 2000, pp. 15–19.

[14] A. R. Michael Dittenbach, Dieter Merkl, "Using growing hierarchical self-organizing maps for document classification," in *ESANN'2000 proceedings: European Symposium on Artificial Neural Networks*, April 2000, pp. 7–12.

[15] S. Robertson and K. S. Jones, "Relevance weighting of search terms," *Journal of the American Society for Information Science*, vol. 27, no. 3, pp. 129–146, 1976.